Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 3 (2025)

Trends and Growth of Productivity of Agricultural Sector usingStatistics Through Machine Learning in Assam

Smrita Borthakur[1], Ranjan Kr. Sahoo[2], Sarat Ch. Kakaty[3], Supahi Mahanta[*]

[1]Research Scholar, Department of Statistics, Central University of Haryana, India ORCID-0000-0001-9734-3795

[2]Department of Statistics, Central University of Haryana, India

[3]Department of Statistics, Dibrugarh University Assam, India

[*]Department of Statistics, Assam Agricultural University Assam, India ORCID-0000-0002-7200-4436

*Corresponding author. E. mail: supahi.mahanta@aau.ac.in

Article History:

***Received:** 02-08-2024*

***Revised:** 23-09-2024*

***Accepted:** 05-10-2024*

Abstract:
The study of trend and growth of agricultural sector by applying statistical techniques through machine learning analysis has become a powerful tool in in the long run. period, In this study Trend Analysis have been done using fitting of curves of the types, Exponential, Modified Exponential, Gompertz and Power Curves by taking secondary data for a period of 1993to 2023. The data have been collected from the official records of Directorate of Economics and Statistics Assam. Trend analysis have been done using Mann-Kendall test and Sen's slope estimator. From the growth models modified exponential is adjudged by higher $R2$ and lowest MSE and lowest AIC with significant t values.

Keywords: Trend, Growth, Exponential, Modified Exponential, Gompertz and Power Curves, Mann- Kendall test and Sen's slope estimator

1. Introduction

Growth is a key process for development. Kumar et al. (2017), Fugile (2018), in their study talked on the deceleration of productivity, highlighting its implications for global food security. Sherman et al. (1950) together with Spurr and Arnold (1948) are the pioneers for showing Simplified Procedures for Fitting Growth Curves for a given set of data. In the realm of agricultural economics, several seminal studies have applied a diverse range of growth curve models to unravel trends in agricultural development. (Aheam et al. (1998), Greeshma et al. (2014), Jaypatre et al. (2010), Sreenivas et al. (2019), Suthar et al. (2024)) conducted pioneering research, employing Exponential, Modified Exponential, Logistic, Gompertz, and Power Curves to analyze agricultural trends comprehensively. Their work not only involved the computation of growth rates and compound growth rates but also provided valuable insights into the dynamics of agricultural area, production, and productivity in terms of GDP. This foundational research laid the groundwork for subsequent investigations, such as Adinata et al. (2022) and Brand et al. (2024), which further delved into the intricacies of agricultural growth for animals using similar methodologies.

Building upon these early works, Williams et al. (2003) explored the implications of growth curve analysis for agricultural policy formulation. Subsequent studies by Rahman et al. (2011) and Li et al. (2010) expanded the scope of inquiry by examining the role of technological advancements in driving agricultural productivity growth, utilizing sophisticated modelling techniques.

In parallel, Mann-Kendall and Sen's slope analysis has emerged as a powerful tool for trend analysis in agricultural economics. Studies by (Binod et al. (2022), Ham et al. (2023) and Chandler et al. (2011)) have utilized this method to discern long-term patterns in agricultural production and productivity, reinforcing traditional growth curve analyses.

Chen et al. (2022) suggested applications machine learning techniques for trend analysis. Furthermore, the integration of exploratory data analysis (EDA) techniques, as demonstrated by Shanmugan et al (2018) and Alston et al. (2019), has enriched our understanding of the underlying drivers of agricultural growth, facilitating the identification of key factors influencing agricultural dynamics.

These studies collectively contribute to a deeper understanding of the complex dynamics shaping agricultural economies worldwide, highlighting the pivotal role of growth curve analysis, Mann-Kendall and Sen's slope analysis, and exploratory data analysis in elucidating trends and informing policy decisions.

2. Data and methodology:

Data collection: The time-series data on fish productivity (in Kg/hectare) in Assam for a period of 30 years from 1993 to 2023 was obtained from the Directorate of Economics and Statistics, Government of Assam for conducting the present study.

The methods and Models we have used here are least squares, Men Kandell Sen's Slope, 3-point sum using Exploratory data analysis in python. The models that have been employed are

(A) Exponential, logistic, Modified Exponential, Gompertz and Power Curve

- y_t=a b^t (1)

Growth rate of the equation is

$$\frac{dy_t}{dt} = a \frac{db^t}{dt}$$

$$= a \log b * b^t$$

$$= \log b * y_t$$

$$\frac{dy_t}{dt} \Big/ y_t = \log b \qquad (2)$$

- $y_t = a\beta^t\gamma^{t^2}$ (3)

Growth rate of the above equation is

$$\frac{dy_t}{dt} = a[\gamma^t \frac{d\beta^{t^2}}{dt} + \beta^t \frac{d\gamma^{t^2}}{dt}]$$

$$= a[\gamma^{t^2} \log\beta\, \beta^t + \beta^t 2\log\gamma\, \gamma^{t^2}]$$

= Y logβ+ Y2logγ

$$\frac{dy_t}{dt} / y_t = (\log\beta + 2\log\gamma) \qquad (4)$$

- y_t=K +ab^t (5)

Growth rate of the above equation is

$$\frac{dy_t}{dt} = = a\log b\, b^t$$

$$= \log b\; a b^t$$

$$\frac{dy_t}{dt} / y_t = \log b (1 - K/y_t) \qquad (6)$$

- y_t=K ab^t (7)

Growth rate of the above equation is

$\frac{dy_t}{dt}$=K $\frac{db^t}{dt}$

=K log a ab^t

=log a y_t

$\frac{dy_t}{dt}/y_t$ = log a (8)

- y_t=A e^{rt} (9)

Growth rate of the above equation is

$\frac{dy_t}{dt}$ = A r e^{rt}

=r y_t

$\frac{dy_t}{dt}/y_t$ = r (10)

- y_t=αt^β (11)

Growth rate of the above equation is

$\frac{dy_t}{dt}$= α $\frac{dt^\beta}{dt}$

= $\alpha\beta t^{\beta-1}$

$\frac{dy_t}{dt}/y_t$ = $\frac{\beta}{t}$ (12)

To determine the magnitude and direction of the trend Mann Kendall and Sen's slope have been used.

(B) Mann-Kendall Test:

Rank-based nonparametric methods provide alternative statistical approaches to the conventional parametric methodology. Recent advancements in the field of individual rank tests have been substantially influenced by the works of Zimmerman and Wolfowitz (1940), Hotelling and Pabst (1936), Kendall (1938), Smirnov (1939), Mann (1945), and Wald and Wolfowitz (1940), despite the older nature of the concept. Savage came across over three thousand publications pertaining to nonparametric procedures, particularly rank-based methods, in 1962. Prior to these advancements, a vast quantity of research papers were published concerning nonparametric methods. However, during that period, studies were concerned about the efficacy of nonparametric methods due to the weak assumptions they required for validity. Historiographic validation of the Wilcoxon test, along with other nonparametric techniques, is attributable to Chernoff and Savage (1958) and Hodge and Lehmann (1956), provided that the normality assumption is satisfied. However, in circumstances where the normality assumption is not met, the procedures might prove to be more impactful and legitimate. Mann-Kendall and Modified Mann-Kendall (Haled and Rao, 1998) are the primary non-parametric techniques utilized in hydroclimatic time series trend analysis (Kendall, 1955; Mann, 1945). With regard to the orientation of the trend, the Mann-Kendall's trend test examines time series data for trends, notwithstanding its non-parametric nature (Kisi et al. 2004, Partal and Kahya, 2006; Yenigun et al., 2008; Hadgu et al., 2013, Alhaji et al. (2018)). Consequently, its vulnerability to anomalous effects is diminished.

The MK test compares the alternative hypothesis, which asserts the presence of a trend in either an upward or downward direction, to the null hypothesis, which states that no trend exists. The MK test statistic (S) is given by-

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \operatorname{sgn}(x_j - x_k)$$

$$\operatorname{sgn}(x_j - x_k) = \begin{cases} +1 & \text{if } (x_j - x_k) > 0 \\ 0 & \text{if } (x_j - x_k) = 0 \\ -1 & \text{if } (x_j - x_k) < 0 \end{cases} \quad (13)$$

$$\text{Var} = [n(n-1)(2n+5) - \sum t(t-1)(2t+5)]/18 \quad (14)$$

The notation $\sum t$ represents the summation of all ties, while t represents the extent of a specific tie. In cases where the sample size surpasses 10, the standard normal variate z is computed using Equation (2) (Douglas et al., 2000).

$$z = \begin{cases} (S-1)/\sqrt{\text{Var}(S)} & \text{if } S>0 \\ 0 & \text{if } S=0 \\ (S+1)/\sqrt{\text{Var}(S)} & \text{if } S<0 \end{cases} \quad (15)$$

For a two-sided test, H_0 should be accepted if $z \leq z_{\alpha/2}$ at the given level of significance. A positive value of S indicates an upward trend and a negative value ofS indicates downward trend.

(C) Sen's Slope:

The magnitude of trend in a time series is generally determined by using regression analysis and Sen's estimator (Sen, 1968). The former one is a parametric test and the later one is non-parametric test. Both the two methods assume a linear trend in the time series. In this method, the slopes (T_i) for all data pairs are first calculated by the following way-

$$T = \frac{x_j - x_k}{j - k} \quad \text{for } i = 1,2, \ldots, N \quad (16)$$

here x_j are the values at time j and x_k are the values at time k (j > k). The median of these N values of T_i is Sen's estimator of slope, which is calculated as

$$L = \begin{cases} T_{\frac{N+1}{2}} & \text{if N is odd} \\ \frac{1}{2}\left(T_{N/2} + T_{\frac{N+2}{2}}\right) & \text{if N is even} \end{cases} \quad (17)$$

A time series is characterized by an upward trend when the value of L is positive, and a downward trend when the value is negative. Checking of trend of productivity of crops was done by trend package included in the software Python

(D) EDA:

EDA is a part of ML technique used to analyze data using both non-visual and visual technique. Tukey (1977) and Mathew et al. (2016) mentioned the graphical and non-graphical methods of Exploratory data analysis for applied sciences for trend analysis. There is a structured approach which is followed Involves thorough analysis of data to understand the current situation. Through EDA we can extract meaningful information based on domain understanding. The codes involved in EDA are

Importing the dataset

```
# To import the data
```

Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 3 (2025)

```
df = pd. read_csv('C:/Users/smrit/Downloads/Cob.csv')
```

For removing outliers

```
def remove_outlier(col):
    sorted(col)
    Q1, Q3= np. Percentile (col, [25,75])
    IQR=Q3-Q1
    lower_range= Q1-(1.5 * IQR)
    upper_range= Q3+(1.5 * IQR)
    return lower_range, upper_range

for column in df. Iloc [ :, 1:7]. columns:
    lr, ur=remove_outlier(df[column])
    df[column]=np. where (df[column]>ur, ur, df[column])
    df[column]=np. where (df[column]<lr, lr, df[column])
```

For standardization

Scale the numeric columns scaled_data = scaler.fit_transform(df[numeric_cols])# Replace the original numeric columns with scaled data

df_scaled = pd. Data Frame (scaled_data, columns=numeric_cols)

3. Results and Discussion:

The Growth and trends of the productivity of the crops of Assam.Growth Curves of productivity of crops of Assam

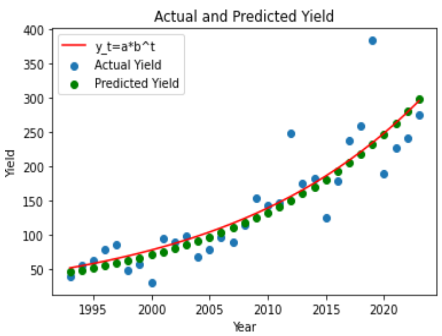

Fig9: Exponential Curve showing the actual, fitted andpredicted lines

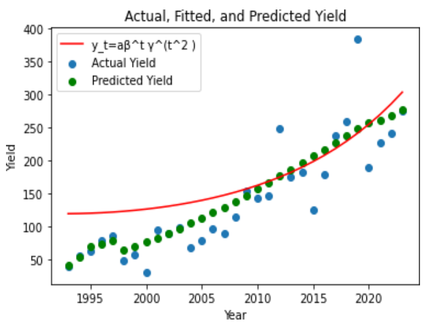

Fig10: Log parabolic Curve showing the actual, fitted and predicted lines

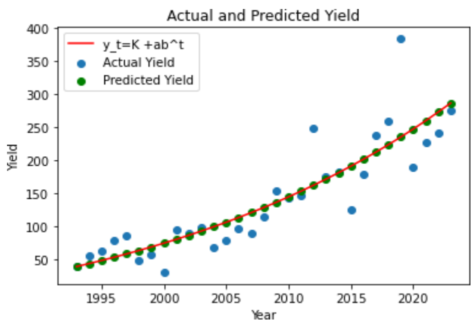

Fig11: Modified Exponential Curve showing the actual, fittedand predicted lines

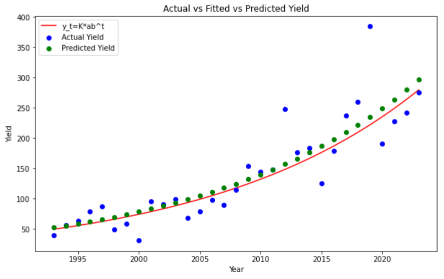

Fig12: Gompertz Curve showing the actual, fitted andpredicted lines

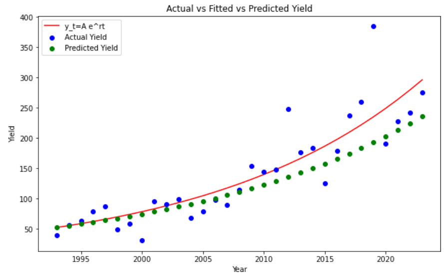

Fig13: Exponential Curve showing the actual, fitted andpredicted lines

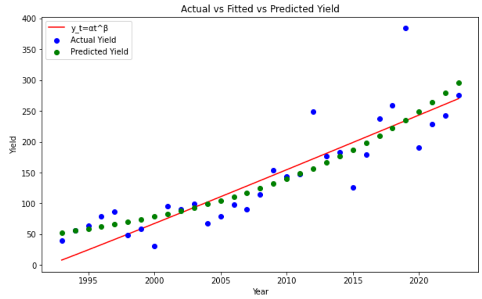

Fig14: Power Curve showing the actual, fitted and predicted lines

Fit Statistics of the models:

Model	R^2	MSE	t_{cal}	Prob	AIC	Instantaneous Growth rate
y_t=a b^t	0.764	1633.68	0.765	1.26e-10	321.27	0.89
y_t =a$\beta^t\gamma^{t^2}$	0.544	3159.93	0.889	2.345e-11	255.81	`0.02
y_t=K +ab^t	0.7701	1595.35	0.877	9.137e-11	234.62	0.93
y_t=K ab^t	0.753	1713.01	0.875	1.268e-10	236.83	0.82
y_t=A e^{rt}	0.764	1633.63	0.875	1.266e-10	236.83	0.056
$y_t=\alpha t^\beta$	0.744	1772.80	0.874	1.2616e-10	235.88	0.066

From the fit Statistics of the models, we can see that y_t=K +ab^t have the highest R square i.e. 0.77 with lowest AIC i.e 234.62 is a good model for growth identification as compared to other models.

Trend identification for productivity of crops using Mann-Kendall andSen's Slope estimator

Crops	p- value of Mann-Kendall test	Significance(α)	Sen's slope
Rice	**0.01**	**	**-12.0012**
Wheat	**0.092**	_	**-1.0078**
Urad	**0.022**	**	**-2.0072**
Tur	**0.001**	**	**-1.0074**
Gram	**0.704**	_	**0.01**
Maize	**0.641**	_	**1.62**
Masoor	**0.289**	_	**2.41**
Moong	**0.289**	_	**-2.22**
Small Millets	**0.442**	_	**0.01**
Other Rabi Pulses	**0.351**	_	**-3.12**
Rapeseed Mustard	**1.62e-01**	**	**-10.001**
Castor seed	**0.721**	_	**-0.01**
Linseed	**1.32e-08**	**	**-1.01**
Sesamum	**0.363**	_	**-0.01**
Niger	**0.547**	_	**0.01**

Jute	**9.36e-05**	**	**0.001**
Mesta	**0.349**	_	**0.002**
Cotton	**1.494e-05**	**	**-0.002**
Sweet Potato	**0.001**	**	**0.005**
Potato	**0.043**	**	**5.008**
Sugarcane	**0.002**	**	**-5.002**
Turmeric	**0.032**	**	**2.001**
Onion	**2.67e-05**	**	**3.007**
Ginger	**1.97e-12**	**	**0.002**
Black paper	**3.95e-06**	**	**-0.004**
Arecanut	**0.0001**	**	**0.002**
Banana	**0.001**	**	**5.002**
Coconut	**0.079**	_	**1.239**
Tapioca	**0.183**	_	**0.003**
Dry Chilies	**0.005**	**	**-0.0071**
Peas and Beans	**0.13**	_	**-5.88**
Tobacco	**9.768e-15**	**	**0.005**

***: 5% level of significance**

****: 1% level of significance**

The results of the analysis show that Jute, Sweet Potato, Potato, Turmeric, Onion, Ginger, Arecanut, Ban ana and Tobacco have a significant increasing trend, as indicated by the positive value of the Sen's slope, which indicates an increasing tendency of productivity by 0.001, 0.001, 5.005, 0.008, 2.001, 3.007, 0.00 2, 0.002, 5.002, 0.005 kg per hectare per year under the mentioned crop cultivation respectively. Gram, Maize, Masoor Small millets, Niger, Mesta, Coconut and Tapioca on the other hand, has an insignificant increasing trend. Wheat, Moong, Other Rabi Pulses, Castor seed, Sesamum, Peas and Beans have decrea sing insignificant trend. Rice, Urad, Tur, Rapeseed Mustard, Linseed, Cotton, Sugarcane, Black Paper an d Dry chilies has a decreasing significant trend with productivity decreased by -12.0012, -2.0072, -1.007 4, -10.001, -1.01, -0.002, -5.002, -0.004, -0.0071 kg per hectare per year, respectively.

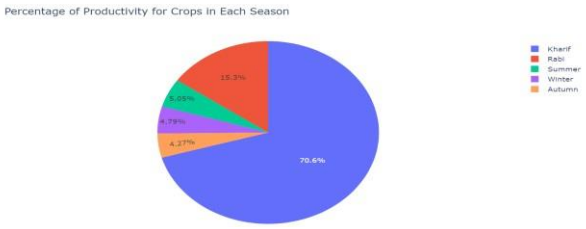

Fig 1 productivity (yield) of crops are more in

kharif followed by rabi, summer, winter, autumn

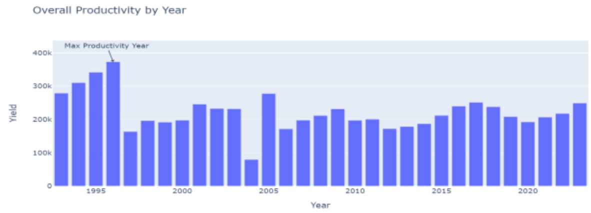

Fig 2 productivity (yield) of crops year wise

which is in mix bag of swings

Conclusion:

The agricultural sector in Assam stands at a critical juncture where leveraging advanced statistical techniques and machine learning algorithms could significantly enhance productivity, leading to substantial economic growth. By analysing historical data and employing predictive models, such as growth curves and other forecasting methods, policymakers, researchers, and stakeholders can gain valuable insights into trends and make informed decisions to propel the agricultural sector forward.

Statistical analysis reveals patterns and trends in crop yields, land usage, weather conditions, and socio-economic factors, providing a foundation for predictive modelling. Statistical techniques with Machine learning algorithms offer powerful tools to forecast agricultural productivity accurately.

Techniques like growth curves, time series analysis, and ensemble methods can capture complex relationships within the data, enabling precise predictions.

Accurate predictions enable policymakers to implement targeted interventions, such as crop selection recommendations, irrigation planning, and pest control strategies, maximizing output while minimizing input costs. Enhanced productivity in the agricultural sector leads to increased income for farmers, improved food security, and overall economic development.

Moreover, by harnessing machine learning for agricultural innovation, Assam can attract investments, foster technological advancements, and establish itself as a leader in sustainable farming practices.

While machine learning holds immense potential, it's crucial to address challenges such as data scarcity, quality issues, and model interpretability.

Collaborative efforts involving government agencies, research institutions, and technology providers are essential to overcome these obstacles and ensure the successful implementation of predictive analytics in agriculture.

In conclusion, the convergence of statistics and machine learning presents a transformative opportunity for the agricultural sector in Assam. By harnessing the power of data-driven insights, policymakers can steer the region towards a path of inclusive economic growth, agricultural sustainability, and prosperity for all stakeholders involved.

References:

[1] Ahearn, M., Yee, J. E., Nehring, B. R. (1998). Agricultural Productivity in the United States. USDA,ERS, Agr. Infor. Bull. No. 740, Jan. 1998.

[2] Adinata, Y., Noor, R. R., Priyanto, R., Cyrilla, L., Sudraja, P. (2022). Comparison of growth curve models for Ongole Grade cattle, Tropical Animal Health and Production, 54:252.

[3] Alhaji, U. U., Yusuf, S. A., Edet, C. O., Oche, C. O., Egbe, E. P. (2018). Trend Analysis of Temperature in Gombe State Using Mann Kendall Trend Test, Journal of Scientific Research and Reports, Vol- 20, Issue-3.

[4] Alston, J. M., & Pardey, P. G. (2019). Transforming Traditional Agriculture Redux. In C. Monga & J.Lin (Eds.), The Oxford Handbook of Structural Transformation. Oxford: Oxford University Press.

[5] Brand, S. T., Westhuyzen, J. P., Hough, W., Zyl, J. (2024). Application of growth models to South African Boer goat castrates and does under feedlot conditions, Tropical Animal Health and Production, 56:178.

[6] Chandler, R., & Scott, M. (2011). Statistical Techniques for trend detection and analysis in the environmental sciences. John Wiley & Sons.

[7] Chen, Z., Qiu, Y., Song, Z. (2022). Short-term stock trends prediction based on sentiment analysisand machine learning, Springer Link, Volume 26, pages 2209–2224.

[8] Chernoff, H., & Savage, I. R. (1958). Asymptotic normality and efficiency of certain nonparametrictest statistics. Annals of Mathematical Statistics, 29, 972–994.

[9] Douglas, E. M., R. M. Vogel, and C. N. Kroll (2000), Trends in floods and low flows in the United States: Impact of spatial correlation, J. Hydrol. Amsterdam, 240, 90 – 105, doi:10.1016/S0022- 1694(00)00336-X.

[10] Fugile K.O. (2018). Is agricultural productivity slowing? Global Food Security 17, 73-83.

[11] GREESHMA. R, BHAVE. M, SUPRIYA K. (2014). GROWTH TRENDS OF SUGARCANE (Saccharum officinarum) CROP IN ANDHRAPRADESH, The J. Res. PJTSAU 42(3)84-86, 201484.

[12] Hodges, J., Lehmann, E. (1956). The efficiency of some Nonparametric Competitors of the t-Test,Annals of Mathematical Statistics, 391-402.

[13] Hotelling, H. and Pabst, M. R.: Rank Correlation and Tests of Significance Involving No Assumption of Normality, The Annals of Mathematical Statistics, 2957508, 29–43, 1936

[14] Haled, K.H., Rao, A.R. (1998). A modified Mann-Kendall trend test for autocorrelated data, Journal of Hydrology, Volume 204, Issues 1–4, 182-196.

[15] Hadgu, G., Kindie T., Girma M., Biley K. (2013). Trend and variability of rainfall in Tigray, northernEthiopia: analysis of meteorological data and farmers' perception, Academic J agri Res (1)6: 088-100

[16] Jaypatre J.S., Patel K.S., Gajbhiye S.B., Awaghad P.R. (2010). Performance of Mango in South Gujrat Region. Agriculture Update, Volume-5, 3&4, 480-482.

[17] Kendall, M. G. (1938). A new measure of rank correlation. Biometrika 30 (1/2), 81–93.

[18] Kendall, M. G., 1955: Rank Correlation Methods. 2d ed. Charles Griffin, 196 pp.

[19] Kisi et al. (2004). Comparison of Mann–Kendall and innovative trend method for water quality parameters of the Kizilirmak River, Turkey, Journal of Hydrology, Volume 513, 26, 362-375.

[20] Kumar V., Patel G.T., Singh D.R., Chand R. (2017). Trend of Area, Production and Productivity of Major Cereal Crops in Context of Food Security: Sikkim, India. International Journal of Engineering Technology Science and Research, ISSN 2394-3386, Volume 4, Issue 9

[21] Li, H., J. Qiu, L. Wang, H. Tang, C. Li, and E. Van Ranst (2010), Modelling impacts of alternative farming management practices on greenhouse gas emissions from a winter wheat-maize rotationsystem in China, Agric. Ecosyst. Environ., 135, 24–33, doi:10.1016/j.agee.2009.08.003.

[22] Mann, H.B. 1945. Non-parametric tests against trend, Econometrica 13:163-171.

[23] Matthieu Komorowski, Dominic C. Marshall, Justin D. Salciccioli and Yves Crutain (2016). Exploratory Data Analysis, https://www.researchgate.net/publication/308007227

[24] Partal T., Kahiya E. (2006). Trend Analysis in Turkish precipitation data Hydr Proc 20: 2011-2026.

[25] REHMAN. F, SAEED. I, SALAM. A (2011). ESTIMATING GROWTH RATES AND DECOMPOSITION ANALYSIS OF AGRICULTURE PRODUCTION IN PAKISTAN: PRE AND POST SAP ANALYSIS, Sarhad J. Agric. Vol.27, No.1, 2011.

[26] Sen PK (1968) Estimates of the regression coefficient based on Kendall's Tau. J Am Stat Assoc 63(324):13791389.

[27] Shanmugan, K. D., Baria, B. P. (2018). Total factor productivity in Indian agriculture: An exploratoryanalysis based on literature review, Agricultural Reviews, DOI:10.18805/ag.R-1754

[28] Sherman. J, Morrison. W. J (1950). Simplified Procedures for Fitting a Gompertz Curve and a Modified Exponential

Curve, Journal of the American Statistical Association, 45(249), 87-97.

[29] Smirnov N. 1939. Ob uklonenijah empiriceskoi krivoi raspredelenija. Recueil Mathema- tique (Matematiceskii Sbornik), NS 6:3–26

[30] Sreenivas. A, Srinivasa. C.D, Supriya. K (2019). Growth Rates, Growth Models and Future Projections of Rice in Three Districts of Northern Telangana Zone, International Journal of CurrentMicrobiology and Applied Sciences, ISSN: 2319-7706 Volume 8 Number 06 (2019.

[31] Spurr, W.A., Arnold, D. R. (1948). A Short-Cut Method of Fitting a Logistic Curve, Journal of the American Statistical Association, 43, 241,127-134.

[32] Suthar B., Pundir R.S., Gundaniya H., Mishra K. (2024). Growth and instability analysis of area, production and yield of groundnut in selected states of India, Environmental Conservation Journal 25(1), 192-198

[33] Tukey J (1977) Exploratory data analysis. Pearson, London

[34] Wald A, Wolfowitz J (1940) On a Test Whether Two Samples are from the Same Population. AnnMath Stat 11:147–162. https://doi.org/10.1214/aoms/1177731909

[35] Williams, M., et al. (2003). Agricultural Policy/Environmental Extender Model Theoretical Documentation., Agri Life Research.

[36] Yenigun K., Gumus V., Bulut H. (2008). Trends in stream flow of the Euphrates basin, Turkey. Water Manag 161: 181-198.

[37] Zimmerman, D. w., & Zumbo, B. D. (1940). The relative power of the WiIcoxon-Mann-Whitney test and Student t-test under simple bounded transformations. Journal of General Psychology. 1I7. 425-436.